The Contribution of Individual Variables to Hotelling's T^2, Wilks' Λ, and R^2

Author(s): Alvin C. Rencher
Source: *Biometrics*, Vol. 49, No. 2 (Jun., 1993), pp. 479-489
Published by: International Biometric Society
Stable URL: http://www.jstor.org/stable/2532560
Accessed: 11-03-2016 08:26 UTC

BIOMETRICS 49, 479–489
June 1993

The Contribution of Individual Variables to Hotelling's T^2, Wilks' Λ, and R^2

Alvin C. Rencher

Department of Statistics, Brigham Young University, Provo, Utah 84602, U.S.A.

SUMMARY

We examine the effect of each variable on the following statistics: the one-sample and two-sample Hotelling's T^2, Wilks' Λ for multivariate analysis of variance, and R^2 in multiple regression. For T^2, the net effect of each variable is an increase in the multivariate statistic, and the particular factors determining the amount of increase are (i) the multiple correlation of the variable with all other variables, and (ii) how well the variable's contribution to falsifying the hypothesis can be linearly predicted from the other variables. The effect of each predictor variable on R^2 is similar to the effect of each variable on T^2. For Wilks' Λ, each variable induces a decrease, due to (i) the F for that variable alone, and (ii) the change in multiple correlation from within-sample to total-sample.

1. Introduction

Both univariate and multivariate approaches have been used to assess the contribution of each variable to a multivariate statistic such as Wilks' Λ. A univariate method focuses on the variable of interest and ignores the other variables, whereas a multivariate technique takes into account the presence of other variables.

Three common univariate approaches are individual F tests, Bonferroni tests, and simultaneous tests. Rencher and Scott (1990) found that if these univariate tests are employed only after rejection by the multivariate test, then the "protected" F test on each variable best preserves the α level.

In biological data, it is often the case that many of the variables are highly correlated. Thus, a multivariate approach would be preferred because of the information it provides about the unique contribution of each variable in the presence of the others. This will typically be different from the contribution of a variable by itself and will depend on which other variables are present.

Familiar multivariate approaches to evaluating the contribution of individual variables to a multivariate test statistic include partial F tests of each variable adjusted for the others, standardized canonical discriminant function coefficients, and correlations between the variables and the discriminant function. These correlations were defined in an attempt to find the contribution of each variable independent of the other variables present. However, Rencher (1988) showed that the correlations have this property because they are proportional to individual (not partial) F tests. Thus, this is a univariate method, not a multivariate one, and we will not discuss it further.

In Sections 2, 3, and 4 we present a detailed breakdown of the effect of each variable. In order to place this in the context of existing methods of assessing the effect of individual variables, we review partial F tests in Section 1.1, the related method of subset selection in Section 1.2, standardized discriminant coefficients in Section 1.3, and some relationships among T^2, Λ, and R^2 in Section 1.4. In Section 1.5 we discuss the relationship between the new approach to be introduced in Sections 2, 3, and 4 and the previous work reviewed in Sections 1.1–1.4.

1.1 *Tests for Additional Information*

We begin with the partial F test for the significance of the change in T^2 due to each variable. This so-called "test for additional information" was originally proposed by Rao (1952, pp. 252–253). For more recent discussions see Rao (1973, pp. 554–555) or Seber (1984, pp. 47–54).

For convenience we denote the variable of interest by z, with the remaining variables designated by $\mathbf{x}' = (x_1, x_2, \ldots, x_p)$. Let $T^2_{x,z}$ be the two-sample T^2 statistic based on $(\mathbf{x}', z)$ with T^2_x similarly

Key words: Effect of individual variables on a multivariate test; Hotelling's T^2; Multiple correlation; Partial F test; Standardized discriminant coefficients; Subset selection; Test for additional information; Wilks' Λ.

based on $\mathbf{x}$. Then we test for significance of the increase from T_x^2 to $T_{x,z}^2$ by

$$F_{z|x} = (n_1 + n_2 - p - 2)\left(\frac{T_{x,z}^2 - T_x^2}{n_1 + n_2 - 2 + T_x^2}\right), \tag{1}$$

which is compared to $F_{\alpha,1,n_1+n_2-p-2}$. Huberty (1976) discussed the equivalence of the test for additional information and multivariate analysis of covariance for two groups. This covariance analogy can also be extended to the test statistic (2) below for more than two groups.

For Wilks' Λ, we test for significance of the decrease from Λ_x to $\Lambda_{x,z}$ by (see Rao, 1973, pp. 547–554)

$$\Lambda_{z|x} = \frac{\Lambda_{x,z}}{\Lambda_x}. \tag{2}$$

The partial Λ in (2) can be transformed to an exact F,

$$F_{z|x} = \frac{1 - \Lambda_{z|x}}{\Lambda_{z|x}} \cdot \frac{\nu_E - p}{\nu_H}, \tag{3}$$

which is compared to F_{α,ν_H,ν_E-p}, where ν_H and ν_E are the degrees of freedom for hypothesis and error, respectively.

In multiple regression, we are interested in the increase in R^2 when a predictor variable z is added to the predictors $\mathbf{x}' = (x_1, x_2, \ldots, x_q)$. If y is regressed on $\mathbf{w}' = (\mathbf{x}', z)$, then the partial F test for the significance of the increase from R_{yx}^2 to R_{yw}^2 is given by (see Seber, 1977, pp. 112–113)

$$F_{z|x} = \frac{(R_{yw}^2 - R_{yx}^2)(n - q - 1)}{1 - R_{yw}^2}, \tag{4}$$

which is compared to $F_{\alpha,1,n-q-1}$.

Fujikoshi (1989) gives a good review of these and other tests for additional information and provides formal statements of each hypothesis as well as alternative test statistics in some cases.

1.2 *Subset Selection*

The test for additional information provided by each variable is often used in selection of variables by a stepwise or similar method. When comparing two or more groups, the experimenter may measure a large number of variables so as to include any that might prove effective in separating the groups. For a given sample size, however, an increase in the number of noninformative variables leads to reduced power of tests and increased variability of estimates. Hence, reducing the number of variables is useful in many cases.

A stepwise procedure typically uses the partial F statistic (3) at each step to determine whether a variable should be added to the subset to be retained (or possibly deleted). This procedure is usually called *stepwise discriminant analysis*, although a more meaningful description might be stepwise MANOVA (or stepwise T^2 in the case of two groups). Good surveys of subset selection applied to group separation are given by McKay and Campbell (1982), Huberty (1975, 1984), Krishnaiah (1982), and Berk (1980).

When the variable that maximizes the partial F is chosen, the statistic no longer has the F distribution. This method of choosing gives an optimistic bias to the procedure if nominal significance levels are used. This bias may lead to including too many variables in the subset or even to selecting an entirely spurious subset. Rencher and Larson (1980) showed that the bias is substantial in many cases, especially where the sample size, relative to the number of variables, is small. Hawkins (1976) suggested that a variable be included in the subset only if its partial F is significant at the $\alpha/(k - p)$ level, where α is the desired level of significance, p is the number of variables already entered into the subset, and k is the total number of variables. As with other Bonferroni-based procedures, this is conservative.

All of the above considerations apply equally well to selecting a subset of predictor variables in multiple regression. Stepwise procedures are generally based on the partial F in (4), in which case there is an optimistic bias in R^2 due to selecting the best variable at each step. This bias has been discussed by Pinault (1988), Flack and Chang (1987), Freedman (1983), McIntyre et al. (1983), Rencher and Pun (1980), Berk (1978), and Pope and Webster (1972). On the other hand, the possibility of deleting a variable that should have been retained is discussed by Mandel (1989).

1.3 *Standardized Discriminant Function Coefficients*

For two groups, the discriminant function was defined by Fisher (1936) as the linear combination $z = \mathbf{a}'\mathbf{x}$ that maximally separates the (transformed) means of the two groups. The resulting coefficient vector is (any multiple of) $\mathbf{a} = \mathbf{S}_{\mathrm{pl}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$, where $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ are the sample mean vectors of the two groups and $\mathbf{S}_{\mathrm{pl}}$ is the pooled sample covariance matrix. For k groups, there are $k - 1$ (canonical) discriminant functions whose coefficients are obtained as eigenvectors of $\mathbf{E}^{-1}\mathbf{H}$, where $\mathbf{E}$ and $\mathbf{H}$ are the error and hypothesis matrices from MANOVA.

For the two-group case, Rencher and Scott (1990) expressed the standardized discriminant function coefficients in a form showing precisely how the influence of each variable on the discriminant function depends on overlap with the other variables. If the variable of interest is denoted by z, with the remaining variables designated by $x_1, x_2, \ldots, x_p$, then the standardized discriminant function coefficient corresponding to z is given by

$$\begin{aligned} a_z &= \frac{D_z - \hat{D}_z}{1 - R^2} \\ &= \frac{(\bar{z}_1 - \bar{z}_2)/s_z - \hat{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)/s_z}{1 - R^2}, \end{aligned} \tag{5}$$

where $\bar{z}_i$ ($i = 1, 2$) is the mean of z for the ith group, $\bar{\mathbf{x}}_i$ is the vector of means of the x's for the ith group, $\hat{\boldsymbol{\beta}}$ is the vector of regression coefficients of z regressed on the x's (corrected for means), and R^2 is the corresponding squared multiple correlation. Thus, the contribution of each variable to the discriminant function is due to its multiple correlation with the other variables and how well its separation of groups can be predicted from the other variables.

From (5) we see that the coefficients are not proportional to univariate t's unless the variables are orthogonal. Rencher and Scott (1990) also showed that the coefficients are not proportional to partial t statistics. From (1), with $n_1 = n_2 = n$, the partial t comparing the two groups on z adjusted for the x's is given by

$$t_{z|x} = \sqrt{[2(n-1) - p]\,\frac{T^2_{x,z} - T^2_x}{2(n-1) + T^2_x}}. \tag{6}$$

The relationship between the standardized discriminant function coefficient a_z and $t_{z|x}$ was given by Rencher and Scott as

$$a_z = \frac{\sqrt{2(n-1) + T^2_x}}{\sqrt{[n(n-1) - np/2](1 - R^2)}}\, t_{z|x}, \tag{7}$$

where R^2 is defined as in (5). Since the coefficient of $t_{z|x}$ is not constant, a_z is not proportional to $t_{z|x}$. For further discussion of interpretation of discriminant functions, see Rencher (1992).

1.4 *Relationship of T^2 to Λ and R^2*

There is a similarity of appearance in the partial F statistics in (1), (3), and (4) because of the direct link between T^2 and Λ and between T^2 and R^2. These relationships not only provide insights into test results, but also constitute a useful computational device, since few general-purpose statistical programs provide for direct computation of T^2.

Two groups can be compared using either T^2 or Λ, which are related as follows (Seber, 1984, p. 44):

$$T^2 = (n_1 + n_2 - 2)\,\frac{1 - \Lambda}{\Lambda}. \tag{8}$$

To relate T^2 to R^2, we define a grouping variable y as $n_2/(n_1 + n_2)$ for each observation in group 1 and as $-n_1/(n_1 + n_2)$ in group 2. Then $\bar{y} = 0$ and the fitted model becomes

$$\hat{y}_i = b_1(x_{i1} - \bar{x}_1) + b_2(x_{i2} - \bar{x}_2) + \cdots + b_p(x_{ip} - \bar{x}_p).$$

Let $\mathbf{b}' = (b_1, b_2, \ldots, b_p)$ be the vector of regression coefficients, R^2 be the squared multiple correlation, and $D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\mathrm{pl}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ be the usual Mahalanobis distance. Then we have the following relationships:

$$D^2 = \frac{(n_1 + n_2)(n_1 + n_2 - 2)R^2}{n_1 n_2 (1 - R^2)}, \tag{9}$$

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} D^2 = (n_1 + n_2 - 2)\,\frac{R^2}{1 - R^2}, \tag{10}$$

$$\mathbf{a} = \mathbf{S}_{\mathrm{pl}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = \left[\frac{(n_1 + n_2)(n_1 + n_2 - 2)}{n_1 n_2} + D^2\right]\mathbf{b}. \tag{11}$$

Fisher (1936) first noted the connection between two-group discriminant analysis and multiple regression. Further insights have been given by Healy (1965) and Flury and Riedwyl (1985).

1.5 *The Effect of Each Variable*

In this paper we extend the discriminant function results of Rencher and Scott (1990), as reviewed in Section 1.3, to obtain a detailed breakdown of the effect each variable has on T^2, Λ, and R^2. There is some similarity in the breakdown for these three statistics, as would be expected from their mutual relationships reviewed in Section 1.4.

We find that the net contribution of a variable to each multivariate statistic hinges on two kinds of interplay with other variables, as well as on the individual effect of the variable. These three factors reveal the exact source of the significance or nonsignificance of the partial F test associated with the change in T^2, Λ, or R^2, as reviewed in Section 1.1. Our focus is on the factors that determine the size of the change due to a variable; the aim is descriptive rather than inferential.

The two factors that determine the effect of a variable of interest, say z, in the presence of other variables involve (i) the relationship of z to the other variables ignoring the test statistic, and (ii) the influence of the other variables on the contribution of z to the test statistic. These two factors are quantified in Sections 2, 3, and 4.

In addition to its use in a follow-up procedure to a multivariate test, the information in the three factors could also be exploited in sifting the variables for purposes of subset selection. As noted in Section 1.2, there are two kinds of errors possible in subset selection: inclusion of spurious variables and exclusion of valid ones. The breakdown of each variable's contribution into three factors may help identify the legitimate variables and discard those that are not useful.

The net effect of any variable on T^2, Λ, or R^2 may be greater or less than we would have expected from its univariate contribution alone. For example, in the case of Hotelling's T^2 for two groups with equal n, the partial t for a variable z as given in (6) can be greater or less than the t for a variable by itself, $t = (\bar{z}_1 - \bar{z}_2)/(s\sqrt{2/n})$. It is intuitively apparent that overlap with other variables can render a variable partially redundant so that its multivariate contribution is less than its univariate contribution, but to visualize how the contribution of a variable can be enhanced in the presence of the others is far more enigmatic. Illustrations of such seemingly paradoxical situations have been provided by Flury (1989) and Hamilton (1987). However, the breakdown in Sections 2, 3, and 4 of each variable's effect will make clear how the contribution of a variable can be magnified in a multivariate setting. In particular, see the closing paragraph of Section 2.2 and Example 2 in Section 5.2.

The researcher may find it helpful to know why a variable contributed more than expected or less than expected. For example, admission to a university or professional school may be based on previous grades and the score on a standardized national test. An applicant for admission to a university with limited enrollment would submit high school grades and a national test score. These scores would be entered into a regression equation to obtain a predicted value of first-year grade point average at the university. It is typically found that the standardized test increases R^2 only slightly above that based on high school grades alone. This small increase in R^2 would be disappointing to admissions officials who had hoped that the national test score might be a more useful predictor than high school grades. The designers of such standardized tests may find it beneficial to know precisely why the test makes such an unexpectedly small contribution relative to high school grades. An examination of the two factors discussed in Section 4 would show where the standardized test lost so much ground, and this precise quantification of its relationship to high school grades may be useful in designing an improved version.

In Sections 2 and 3 we consider the effect of each variable on the one- and two-sample T^2 and the Wilks' Λ for one-way MANOVA. In Section 4 we discuss each predictor variable's effect on R^2 in a multiple regression analysis. In the proofs we use the inverse and determinant of a partitioned matrix as first obtained by Schur (1917).

2. Effect of Each Variable on Hotelling's T^2

2.1 *One-Sample Case*

Let $\bar{\mathbf{x}}$ and $\mathbf{S}_{xx}$ be the sample mean vector and covariance matrix of n observation vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, where each $\mathbf{x}_i$ consists of measurements on p variables. Under appropriate assumptions, the hypothesis H_0: $\boldsymbol{\mu}_x = \boldsymbol{\mu}_{x0}$ can be tested by Hotelling's T^2 statistic,

$$T_x^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_{x0})'\mathbf{S}_{xx}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_{x0}).$$

This test is seldom of interest, but analogous one-sample tests are used in paired comparisons, profile analysis, repeated measures, growth curves, etc.

For convenience we designate the variable of interest by z, with the other variables denoted by $\mathbf{x}' = (x_1, x_2, \ldots, x_p)$. In some applications, z will be a new variable added to the x's. Typically, there are no additional variables available and we are interested in the effect of each present variable above and beyond the others; i.e., z represents any one of the $p + 1$ variables.

We denote the hypothesized value of z by μ_{z0}, the sample mean and variance of z by $\bar{z}$ and s_z^2, and the vector of sample covariances of z with the x's by $\mathbf{s}_{zx}$. In partitioned form, the sample covariance matrix is

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{xx} & \mathbf{s}_{zx} \\ \mathbf{s}_{zx}' & s_z^2 \end{pmatrix}. \tag{12}$$

The inverse of $\mathbf{S}$ can be written in the form

$$\mathbf{S}^{-1} = \frac{1}{s_{z.x}^2} \begin{pmatrix} s_{z.x}^2 \mathbf{S}_{xx}^{-1} + \mathbf{S}_{xx}^{-1}\mathbf{s}_{zx}\mathbf{s}_{zx}'\mathbf{S}_{xx}^{-1} & -\mathbf{S}_{xx}^{-1}\mathbf{s}_{zx} \\ -\mathbf{s}_{zx}'\mathbf{S}_{xx}^{-1} & 1 \end{pmatrix}, \tag{13}$$

where $s_{z.x}^2 = s_z^2 - \mathbf{s}_{zx}'\mathbf{S}_{xx}^{-1}\mathbf{s}_{zx}$. For the $p + 1$ variables $(\mathbf{x}', z)$, T^2 becomes

$$\begin{aligned} T_{x,z}^2 &= n \begin{pmatrix} \bar{\mathbf{x}} - \boldsymbol{\mu}_{x0} \\ \bar{z} - \mu_{z0} \end{pmatrix}' \mathbf{S}^{-1} \begin{pmatrix} \bar{\mathbf{x}} - \boldsymbol{\mu}_{x0} \\ \bar{z} - \mu_{z0} \end{pmatrix} \\ &= n(\bar{\mathbf{x}} - \boldsymbol{\mu}_{x0})'\mathbf{S}_{xx}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_{x0}) + \frac{n}{s_{z.x}^2}[\mathbf{s}_{zx}'\mathbf{S}_{xx}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_{x0}) - (\bar{z} - \mu_{z0})]^2 \\ &= T_x^2 + \frac{n[\hat{\boldsymbol{\beta}}'(\bar{\mathbf{x}} - \boldsymbol{\mu}_{x0}) - (\bar{z} - \mu_{z0})]^2}{s_z^2(1 - R^2)}, \end{aligned} \tag{14}$$

where $\hat{\boldsymbol{\beta}} = \mathbf{S}_{xx}^{-1}\mathbf{s}_{zx}$ is the vector of regression coefficients of z regressed on the x's (corrected for their means) and $R^2 = \mathbf{s}_{zx}'\mathbf{S}_{xx}^{-1}\mathbf{s}_{zx}/s_z^2$ is the corresponding squared multiple correlation. We note, incidentally, that (14) provides verification that the addition of a variable can only increase T^2.

We can write (14) in the simpler form

$$\begin{aligned} T_{x,z}^2 &= T_x^2 + \left[\frac{\hat{\boldsymbol{\beta}}'(\bar{\mathbf{x}} - \boldsymbol{\mu}_{x0})}{s_z/\sqrt{n}} - \frac{\bar{z} - \mu_{z0}}{s_z/\sqrt{n}}\right]^2 \Big/ (1 - R^2) \\ &= T_x^2 + \frac{(\hat{t}_z - t_z)^2}{1 - R^2}, \end{aligned} \tag{15}$$

where $\hat{t}_z = \hat{\boldsymbol{\beta}}'(\bar{\mathbf{x}} - \boldsymbol{\mu}_{x0})/(s_z/\sqrt{n})$.

The three factors determining the contribution of z to T^2 are conveniently exhibited in (15). The individual effect of z, ignoring the other variables, is given in $t_z = (\bar{z} - \mu_{z0})/(s/\sqrt{n})$. The two factors that reveal the effect of z in the presence of the x's are $\hat{t}_z$ and R^2. In R^2 we have the relationship of z to the x's disregarding the T^2 test, whereas $\hat{t}_z$ is a "predicted" value of t_z showing the influence of the x's on z as it contributes to T^2.

When t_z and $\hat{t}_z$ are of the same sign, the contribution of z will be important if t_z differs substantially from $\hat{t}_z$ or if R^2 is large. In fact, if R^2 is close to 1, z will be a large contributor to T^2 even if $\hat{t}_z$ and t_z are nearly equal. The contribution of z will be negligible if t_z is close to $\hat{t}_z$ and R^2 is not large, in which case most of the evidence $\bar{z}$ provides against the hypothesis is predictable from $\bar{\mathbf{x}}$. On the other hand, if t_z and $\hat{t}_z$ are of opposite sign, their effect is cumulative and z will have a greater impact. Note that if z were orthogonal to the x's ($\hat{\boldsymbol{\beta}} = \mathbf{0}$ and $R^2 = 0$), its contribution to T_x^2 would reduce to t_z^2.

It is not surprising that the increase in T^2 depends on $\hat{t}_z$ and t_z in the form $(\hat{t}_z - t_z)^2$, but we might have expected an inverse relationship to R^2 rather than to $1 - R^2$. It is clear in (14) or (15), however, that the larger the value of R^2, the larger the increase in T^2. Perhaps we can draw an analogy to simple linear regression, where a given difference between y and $\hat{y}$ is more important if r^2 is larger.

2.2 *Two-Sample Case*

In the two-sample case, the T^2 statistic based on $(\mathbf{x}', z)$ can be written in a form analogous to (14):

$$\begin{aligned} T_{x,z}^2 &= \frac{n_1 n_2}{n_1 + n_2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pl}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) + \frac{n_1 n_2}{n_1 + n_2} \frac{[\mathbf{s}_{zx}'\mathbf{S}_{\text{pl}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - (\bar{z}_1 - \bar{z}_2)]^2}{s_z^2(1 - R^2)} \\ &= T_x^2 + \frac{n_1 n_2}{n_1 + n_2} \frac{[\hat{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - (\bar{z}_1 - \bar{z}_2)]^2}{s_z^2(1 - R^2)}, \end{aligned} \tag{16}$$

where $\mathbf{S}_{\text{pl}}$ is the pooled sample covariance matrix of the x's, $\mathbf{s}_{zx}$ is the vector of pooled covariances of z with the x's, s_z^2 is the pooled variance of z, R^2 is the (within-sample) squared multiple correlation of z regressed on the x's, and $\hat{\boldsymbol{\beta}} = \mathbf{S}_{\text{pl}}^{-1}\mathbf{s}_{zx}$ is the corresponding regression coefficient vector. Note the close resemblance of (16) to (5); the contribution of a variable to T^2 involves the same factors as the breakdown of its standardized discriminant function coefficient.

We can write (16) in the same form as (15):

$$T_{x,z}^2 = T_x^2 + \left[\frac{\hat{\boldsymbol{\beta}}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{s_z[(n_1 + n_2)/(n_1 n_2)]^{1/2}} - \frac{\bar{z}_1 - \bar{z}_2}{s_z[(n_1 + n_2)/(n_1 n_2)]^{1/2}}\right]^2 \Big/ (1 - R^2)$$

$$= T_x^2 + \frac{(\hat{t}_z - t_z)^2}{1 - R^2}. \tag{17}$$

The contribution of z to T^2 depends on the same factors as does the one-sample case, and the discussion in Section 2.1 will not be repeated here.

We can see in (16) and (17) the answer to the question raised in Section 1.5 concerning how a variable can contribute more in the presence of other variables than it does by itself. This can happen in various ways: (i) the "predicted" group separation from the other variables, $\hat{t}_z$, may be much greater than the group separation for the variable by itself, t_z, (ii) t_z and $\hat{t}_z$ may be of opposite sign, and (iii) there may be a very high squared multiple correlation, R^2, with the other variables.

3. Effect of Each Variable on Wilks' Λ

In this section we obtain the constituent factors that determine the decrease in Wilks' Λ due to a given variable. The discussion is in terms of the one-way MANOVA model, but the results are directly applicable to any model for which hypothesis and error matrices can be identified.

Let z be the variable of interest, with the remaining variables designated as $\mathbf{x}' = (x_1, x_2, \ldots, x_p)$. The usual hypothesis and error matrices of one-way MANOVA can be written in partitioned form as

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{xx} & \mathbf{h}_{zx} \\ \mathbf{h}_{zx}' & h_{zz} \end{pmatrix} \quad \text{and} \quad \mathbf{E} = \begin{pmatrix} \mathbf{E}_{xx} & \mathbf{e}_{zx} \\ \mathbf{e}_{zx}' & e_{zz} \end{pmatrix}.$$

With this partitioning, Wilks' Λ for the x's and z becomes

$$\Lambda_{x,z} = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|}$$

$$= \frac{|\mathbf{E}_{xx}|(e_{zz} - \mathbf{e}_{zx}'\mathbf{E}_{xx}^{-1}\mathbf{e}_{zx})}{|\mathbf{E}_{xx} + \mathbf{H}_{xx}|[e_{zz} + h_{zz} - (\mathbf{e}_{zx} + \mathbf{h}_{zx})'(\mathbf{E}_{xx} + \mathbf{H}_{xx})^{-1}(\mathbf{e}_{zx} + \mathbf{h}_{zx})]}$$

$$= \Lambda_x \frac{e_{zz}(1 - R_e^2)}{(e_{zz} + h_{zz})(1 - R_{e+h}^2)}$$

$$= \Lambda_x \frac{(1 - R_e^2)}{(1 + cF_z)(1 - R_{e+h}^2)}, \tag{18}$$

where $\Lambda_x = |\mathbf{E}_{xx}|/|\mathbf{E}_{xx} + \mathbf{H}_{xx}|$ is Wilks' Λ based only on the x's, F_z is the usual F statistic comparing the groups using z alone, c is a ratio of degrees of freedom, R_e^2 is the within-groups squared multiple correlation of z regressed on the x's, and R_{e+h}^2 is the corresponding squared multiple correlation ignoring groups.

Thus, the reduction from Λ_x to $\Lambda_{x,z}$ will be important if F_z is relatively large and if z is substantially more correlated with the x's after fitting the MANOVA model than before. Note that if z is orthogonal to the x's, it reduces Λ_x by a factor $1/(1 + cF_z)$.

4. Effect of Each Variable on R^2 in Multiple Regression

Let y be regressed on the predictor variables $\mathbf{w}' = (x_1, x_2, \ldots, x_q, z) = (\mathbf{x}', z)$, where z is the variable whose effect on y we wish to examine. The covariance matrix involving y and $\mathbf{w}$ can be expressed in the form

$$\mathbf{S} = \begin{pmatrix} s_y^2 & \mathbf{s}_{yw}' \\ \mathbf{s}_{yw} & \mathbf{S}_{ww} \end{pmatrix},$$

where $\mathbf{s}_{yw}$ and $\mathbf{S}_{ww}$ can be further partitioned as

$$\mathbf{s}_{yw} = \begin{pmatrix} \mathbf{s}_{yx} \\ s_{yz} \end{pmatrix} \quad \text{and} \quad \mathbf{S}_{ww} = \begin{pmatrix} \mathbf{S}_{xx} & \mathbf{s}_{zx} \\ \mathbf{s}_{zx}' & s_z^2 \end{pmatrix}. \tag{19}$$

Then the squared multiple correlation of y regressed on $\mathbf{w}$ can be written as

$$R^2_{yw} = \frac{\mathbf{s}'_{yw}\mathbf{S}^{-1}_{ww}\mathbf{s}_{yw}}{s^2_y}. \tag{20}$$

The matrix $\mathbf{S}_{ww}$ as partitioned in (19) has the same form as (12) and its inverse is therefore given by (13). Using this form of $\mathbf{S}^{-1}_{ww}$ in the numerator of (20), we obtain

$$\begin{aligned}\mathbf{s}'_{yw}\mathbf{S}^{-1}_{ww}\mathbf{s}_{yw} &= \frac{1}{s^2_{z.x}}(s^2_{z.x}\mathbf{s}'_{yx}\mathbf{S}^{-1}_{xx}\mathbf{s}_{yx} + \mathbf{s}'_{yx}\mathbf{S}^{-1}_{xx}\mathbf{s}_{zx}\mathbf{s}'_{zx}\mathbf{S}^{-1}_{xx}\mathbf{s}_{yx} - s_{yz}\mathbf{s}'_{zx}\mathbf{S}^{-1}_{xx}\mathbf{s}_{yx} - s_{yz}\mathbf{s}'_{yx}\mathbf{S}^{-1}_{xx}\mathbf{s}_{zx} + s^2_{yz})\\ &= \frac{1}{s^2_{z.x}}[s^2_{z.x}s^2_yR^2_{yx} + (\hat{\boldsymbol{\beta}}'_{zx}\mathbf{s}_{yx} - s_{yz})^2],\end{aligned}$$

and (20) becomes

$$R^2_{yw} = R^2_{yx} + \frac{(\hat{\boldsymbol{\beta}}'_{zx}\mathbf{s}_{yx} - s_{yz})^2}{s^2_ys^2_z(1 - R^2_{zx})}, \tag{21}$$

where $\hat{\boldsymbol{\beta}}_{zx} = \mathbf{S}^{-1}_{xx}\mathbf{s}_{zx}$ is the vector of regression coefficients of z regressed on the x's, and $R^2_{zx} = \mathbf{s}'_{zx}\mathbf{S}^{-1}_{xx}\mathbf{s}_{zx}/s^2_z$ is the corresponding squared multiple correlation. We can simplify (21) by using standardized regression coefficients (beta weights), $\hat{\boldsymbol{\beta}}^*_{zx}$:

$$R^2_{yw} = R^2_{yx} + \frac{(\hat{\boldsymbol{\beta}}^{*\prime}_{zx}\mathbf{r}_{yx} - r_{yz})^2}{1 - R^2_{zx}} \tag{22}$$

$$= R^2_{yx} + \frac{(\hat{r}_{yz} - r_{yz})^2}{1 - R^2_{zx}}, \tag{23}$$

where $\mathbf{r}_{yx}$ is the vector of correlations between y and the x's, and $\hat{r}_{yz} = \hat{\boldsymbol{\beta}}^{*\prime}_{zx}\mathbf{r}_{yx}$. Thus, the contribution of z to R^2_{yw} depends on the difference between r_{yz} and $\hat{r}_{yz}$, and on the squared multiple correlation between z and the x's, R^2_{zx}.

Two other well-known properties of R^2 that are verified incidentally in (23) are (a) R^2 can only increase with an additional variable, and (b) if z were orthogonal to the x's, it would add r^2_{yz} to R^2_{yx}; that is, unless the predictor variables are orthogonal, R^2 cannot be partitioned into components uniquely attributable to each variable. It is clear in (23) that the contribution of z can exceed r^2_{yz} as noted by Hamilton (1987).

In (22) and (23) we can now see the specific information available to the designer of the standardized test in the illustration in Section 1.5. In this case, y is the first-year grade point average at the university, z is the score on the standardized test, and $x_1, x_2, \ldots, x_q$ represent high school grades in various subject areas. The increase in R^2 effected by z is $R^2_{yw} - R^2_{yx}$, which is now broken down into $(\hat{r}_{yz} - r_{yz})^2/(1 - R^2_{zx})$. This breakdown would show precisely why z adds so little to R^2. If $\hat{r}_{yz} = \hat{\boldsymbol{\beta}}^{*\prime}\mathbf{r}_{yx}$ is close to r_{yz}, we would examine the coefficients in $\hat{\boldsymbol{\beta}}^{*\prime}$ to determine which of the r_{yx_i} in $\mathbf{r}_{yx}$ are contributing most. An attempt could then be made to redesign the questions so as to reduce these particular r_{yx_i}'s. Another possible way to boost the contribution of z to R^2_{yw} would be to increase R^2_{zx}, that is, to design the questions in the standardized test to be more correlated with high school grades.

5. Examples

We present three examples illustrating the contribution of variables to T^2, R^2, and Λ. For each variable we give the factors that determine the change in the statistic and, for completeness, also show the test of significance of the change due to the variable as reviewed in Section 1.1. The examples illustrate some ways the contribution of a variable may depend on its relationship with the other variables present.

5.1 *Example* 1: *Hotelling's* T^2

Four enzymes were measured in a search for a screening procedure to detect carriers of Duchenne muscular dystrophy, a disease transmitted from female carriers to some of their male offspring (Andrews and Herzberg, 1985, pp. 223–228). The overall $T^2_{x.z}$ statistic comparing carriers and noncarriers on all four variables was 71.70 ($P = 8.9 \times 10^{-10}$). From (17), the contribution to T^2 due to a variable z is given by $T^2_{x.z} - T^2_x = (\hat{t}_z - t_z)^2/(1 - R^2)$, where z represents any one of x_1, x_2, x_3, or x_4, and x indicates the other three variables. The values of $T^2_{x.z} - T^2_x$, $\hat{t}_z$, t_z, and R^2 are given below for each variable:

z	$\hat{t}_z$	t_z	R^2	$T^2_{x,z} - T^2_x$	P-value
x_1	−3.13	−4.37	.39	2.52	.28
x_2	−.35	−5.40	.01	25.70	.0003
x_3	−3.94	−4.03	.47	.02	.93
x_4	−2.43	−6.44	.25	21.59	.00096

The P-value is from the partial F statistic in equation (1) for the test of significance of the increase in T^2 attributable to each variable (test for additional information).

These variables all have similar individual t_z values but exhibit enormous differences in their contribution to T^2. Without the new information in the columns headed by $\hat{t}_z$ and R^2, the researcher does not know why a variable makes a large or small addition to T^2.

The first and third variables increase T^2 by only 2.52 and .02, respectively, despite the significant individual t values of these two variables, −4.37 and −4.03 (P = .0004 and .017, respectively). In these two cases, $\hat{t}_z$ is so close to t_z that the effect of the variable on T^2 is effectively nullified and the variable becomes almost totally redundant. These two variables have the largest values of R^2, but they are not large enough to offset the closeness of $\hat{t}_z$ to t_z. On the other hand, the second and fourth variables make a large contribution to T^2 because their values of t_z are not well predicted by the other variables.

The specific information about why x_3, for example, contributes almost nothing to T^2 in spite of its large t value may be useful in a search for a variable to replace it. The researcher may wish to find variables that significantly boost the T^2 value in order to have available a more sensitive screening procedure. A preliminary examination may reveal that certain types of variables have favorable patterns on $\hat{t}_z$ or R^2, and the search can be narrowed.

5.2 *Example* 2: R^2

Six hematology variables were measured on 103 workers in an automobile assembly plant (Royston, 1983). For illustrative purposes, we regress one of the variables (lymphocyte count) on the other five. The overall R^2_{yw} for y with the five variables was .92 (F = 101.7, $P \approx 0$). From (23), the increase in R^2 due to a variable z has the breakdown $R^2_{yw} - R^2_{yx} = (\hat{r}_{yz} - r_{yz})^2/(1 - R^2_{zx})$, where z represents any one of $x_1, x_2, \ldots, x_5$ and x represents the other four variables. The values of $\hat{r}_{yz}$, r_{yz}, R^2_{zx}, and $R^2_{yw} - R^2_{yx}$ are given below for each variable:

z	$\hat{r}_{yz}$	r_{yz}	R^2_{zx}	$R^2_{yw} - R^2_{yx}$	P-value
x_1	.02	−.01	.02	.001	.47
x_2	.34	.24	.14	.01	.02
x_3	.08	.78	.44	.85	0
x_4	.46	.02	.37	.32	6.9×10^{-17}
x_5	.10	.12	.04	.0004	.63

The P-value is from the partial F test (4) for the significance of the increase in R^2 due to each variable.

An interesting variable here is x_4, with r_{yz} = .02. Despite this small individual correlation with y, x_4 contributes a great deal more to R^2_{yw} than x_2, whose r_{yz} is .24, because $\hat{r}_{yz} - r_{yz}$ and R^2_{zx} are much greater for x_4 than for x_2. This illustrates the discussion in Section 1.5 concerning how the contribution of a variable can be augmented in the presence of other variables.

The difference between the two major contributors, x_3 and x_4, may be very revealing to the researcher. The contribution of x_3 to R^2_{yw} is due mostly to its own correlation with y, whereas all of the effect of x_4 comes from the other variables.

5.3 *Example* 3: *Wilks'* Λ

The data for this example were collected by G. Rex Bryce and Ruel M. Barker (personal communication) as part of a preliminary investigation of a possible link between football helmet design and neck injuries. Six head measurements were made on three groups: high school football players, college football players, and non-football players. There were 30 observations in each group.

In a MANOVA comparison of the three groups, we obtain $\Lambda_{x,z}$ = .31, which is highly significant, as seen by comparison to the critical value, $\Lambda_{.05,6,2,87}$ = .778. From (18), the reduction in Λ due to a variable z has the breakdown $\Lambda_{x,z}/\Lambda_x = (1 - R^2_e)/(1 - R^2_{e+h})(1 + cF_z)$, where z represents any one of $x_1, x_2, \ldots, x_6$ and x represents the other five variables. The values of R^2_e, R^2_{e+h}, F_z, and $\Lambda_{x,z}/\Lambda_x$ are given below for each variable:

z	R_e^2	R_{e+h}^2	F_z	$\Lambda_{x,z}/\Lambda_x$	P-value
x_1	.44	.35	2.6	.81	.0002
x_2	.73	.77	6.2	.99	.9997
x_3	.65	.66	1.7	.99	.999
x_4	.29	.58	58.2	.73	2.6×10^{-6}
x_5	.22	.44	22.4	.91	.02
x_6	.39	.35	4.5	.84	.0009

The P-value is from the partial F test (3) for the significance of the reduction in Λ due to each variable.

Only two variables, x_1 and x_6, have $R_e^2 > R_{e+h}^2$, which enhances the contribution of a variable. Thus, x_1 has a univariate F_z of only 2.6 ($P = .08$), but with $R_e^2 = .44$ and $R_{e+h}^2 = .35$, it ends up with $\Lambda_{x,z}/\Lambda_x = .81$, and reduces $\Lambda_{x,z}$ by more than any other variable except x_4. In contrast, x_2, with an individual F_z of 6.2 ($P = .003$), provides essentially no reduction in $\Lambda_{x,z}$ because it has $R_e^2 < R_{e+h}^2$. On the other hand, x_4 and x_5 have such large univariate F's, 58.2 and 22.4 ($P = 9.7 \times 10^{-17}$ and 1.4×10^{-8}), that the very unfavorable ratio of $1 - R_e^2$ to $1 - R_{e+h}^2$ in both cases was not sufficient to prevent them from making a significant reduction in $\Lambda_{x,z}$.

In this case, x_4, the variable with the greatest multivariate contribution, also has the greatest univariate effect, while x_1, the variable with the next largest multivariate contribution, has an insignificant individual effect. This can be seen by comparing the univariate F_z and the partial F test. But the researcher now has the additional information that the effect of x_4 was greatly dampened by R_e^2 being so much smaller than R_{e+h}^2. A similar problem with x_5 renders it almost insignificant multivariately with a P-value for the partial F test of only .02. Perhaps there is something characteristic of these two variables that would give the researcher a clue as to why their effectiveness is so severely impaired in the presence of the other variables. This information may be useful in redesigning these measurements to increase their effectiveness.

6. Discussion

The effect of each variable on T^2 or R^2 depends on its multiple correlation with the other variables and the predictability of its individual contribution to falsifying the hypothesis. The effect of each variable on Wilks' Λ depends on how well that variable separates the groups by itself and on the improvement in its multiple correlation with the other variables after fitting the MANOVA model.

In the examples, we have illustrated the use of these factors as a follow-up analysis to a partial F test. The factors that determine the significance or nonsignificance of the contribution of each variable can be clearly seen.

It has long been known from applications that the contribution of a variable can be either dampened or enhanced in the presence of other variables. The picture is now complete as to how this takes place. The puzzling effect of the other variables is explicated by the breakdown of the multivariate contribution of a variable in Sections 2, 3, and 4. With this added information the researcher has guidelines for designing or screening new variables that will be more effective.

For example, suppose a researcher is comparing a group of subjects exhibiting high blood cholesterol with a group who have low cholesterol levels. Some variables that may separate these two groups are level of dietary animal fat, level of dietary fiber, level of physical activity, and a measure of hereditary predisposition. The researcher may be especially interested in the interaction of these variables as they contribute jointly to group separation. The interaction of a variable with the others is shown precisely in (16). We can examine $\hat{\boldsymbol{\beta}}$ for an indication of how much each of the other variables influences the variable in question. The overall relationship with the other variables, as measured by R^2, also affects the contribution of a variable.

Another application of the breakdown of the contribution of each variable is in stepwise selection of variables. If redundant variables tend to show a different pattern in the three components of their breakdown than the pattern demonstrated by valid variables, then a more meaningful selection can be made. A selection approach given by Beale, Kendall, and Mann (1967) suggests the elimination of variables that are highly correlated with the variables to be retained in the subset. This cannot be recommended because in (17) and (23) we see that the contribution of a variable actually increases as R^2 increases. This is also true in (18) for R_{e+h}^2, which is based on the total sample ignoring groups. In addition, the breakdown in (17), (18), and (23) shows that two other factors besides the multiple correlation are involved in the contribution of a variable to the test statistic.

ACKNOWLEDGEMENTS

I wish to thank the editor, an associate editor, and a referee for constructive comments that were very helpful in clarifying the exposition.

RÉSUMÉ

Nous examinons l'effet de chaque variable sur les statistiques suivantes: le T^2 de Hotelling pour un et deux échantillons, le Λ de Wilks pour l'analyse de variance multivariable, et le R^2 pour la régression multiple. En ce qui concerne le T^2, l'effet net de chaque variable est un accroissement de la statistique multivariable. Les facteurs particuliers déterminant la valeur de l'accroissement sont (i) le coefficient de corrélation multiple de la variable avec toutes les autres variables, et (ii) comment la contribution de la variable à la fausseté de l'hypothèse peut être linéairement prédite par les autres variables. L'effet de chaque variable prédictrice sur le R^2 est semblable à l'effet de chaque variable sur le T^2. Pour le Λ de Wilks, chaque variable induit une diminution, due à (i) le F de chaque variable prise séparément des autres, and (ii) le changement de la corrélation multiple intra à la corrélation multiple totale.

REFERENCES

Andrews, D. F. and Herzberg, A. M. (1985). *Data*. New York: Springer-Verlag.

Beale, E. M. L., Kendall, M. G., and Mann, D. W. (1967). The discarding of variables in multivariate analysis. *Biometrika* **54,** 357–366.

Berk, K. N. (1978). Comparing subset regression procedures. *Technometrics* **20,** 1–6.

Berk, K. N. (1980). Forward and backward stepping in variable selection. *Journal of Statistical Computation and Simulation* **10,** 177–185.

Fisher, R. A. (1936). The use of multiple measurement in taxonomic problems. *Annals of Eugenics* **7,** 179–188.

Flack, V. F. and Chang, P. C. (1987). Frequency of selecting noise variables in subset regression analysis: A simulation study. *The American Statistician* **41,** 84–86.

Flury, B. W. (1989). Understanding partial statistics and redundancy of variables in regression and discriminant analysis. *The American Statistician* **43,** 27–31.

Flury, B. W. and Riedwyl, H. (1985). T^2 tests, the linear two-group discriminant function, and their computation by linear regression. *The American Statistician* **39,** 20–25.

Freedman, D. A. (1983). A note on screening regression equations. *The American Statistician* **37,** 152–155.

Fujikoshi, Y. (1989). Tests for redundancy of some variables in multivariate analysis. In *Statistical Data Analysis and Inference*, Y. Dodge (ed.). Amsterdam: Elsevier Science Publishers B. V. (North-Holland).

Hamilton, D. (1987). Sometimes $R^2 > r^2_{yx_1} + r^2_{yx_2}$. *The American Statistician* **41,** 129–132.

Hawkins, D. M. (1976). The subset problem in multivariate analysis of variance. *Journal of the Royal Statistical Society, Series B* **38,** 132–139.

Healy, M. J. R. (1965). Computing a discriminant function from within-sample dispersions. *Biometrics* **21,** 1011–1012.

Huberty, C. J. (1975). Discriminant analysis. *Review of Educational Research* **45,** 543–598.

Huberty, C. J. (1976). Covariance and discriminant analysis. *Journal of Experimental Education* **44,** 16–19.

Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. *Psychological Bulletin* **95,** 156–171.

Krishnaiah, P. R. (1982). Selection of variables in discriminant analysis. In *Handbook of Statistics, Volume* 2, P. R. Krishnaiah and L. N. Kanal (eds), 883–892. Amsterdam: North-Holland.

Mandel, J. (1989). Some thoughts on variable-selection in multiple regression. *Journal of Quality Technology* **21,** 2–6.

McIntyre, S. H., Montgomery, D. B., Srinivasan, V., and Weitz, B. A. (1983). Evaluating the statistical significance of models developed by stepwise regression. *Journal of Marketing Research* **20,** 1–11.

McKay, R. J. and Campbell, N. A. (1982). Variable selection techniques in discriminant analysis. I: Description. *British Journal of Mathematical and Statistical Psychology* **35,** 1–29.

Pinault, S. C. (1988). An analysis of subset regression for orthogonal designs. *The American Statistician* **42,** 275–277.

Pope, P. T. and Webster, J. T. (1972). The use of an F-statistic in stepwise regression procedures. *Technometrics* **14,** 327–340.

Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research.* New York: Wiley.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. New York: Wiley.

Rencher, A. C. (1988). On the use of correlations to interpret canonical functions. *Biometrika* **75,** 363–365.

Rencher, A. C. (1992). Interpretation of canonical discriminant functions, canonical variates, and principal components. *The American Statistician* **46,** 217–225.

Rencher, A. C. and Larson, S. F. (1980). Bias in Wilks' Λ in stepwise discriminant analysis. *Technometrics* **22,** 349–356.

Rencher, A. C. and Pun, F. C. (1980). Inflation of R^2 in best subset regression. *Technometrics* **22,** 49–53.

Rencher, A. C. and Scott, D. T. (1990). Assessing the contribution of individual variables following rejection of a multivariate hypothesis. *Communications in Statistics—Simulation and Computation* **19,** 535–553.

Royston, J. P. (1983). Some techniques for assessing multivariate normality based on the Shapiro–Wilk *W*. *Applied Statistics* **32,** 121–133.

Schur, I. (1917). Ueber Potenzreihen die im Innern des Einheitskreises Beschraenkt Sind. *Journal fuer Reine und Angewandte Mathematik* **147,** 205–232.

Seber, G. A. F. (1977). *Linear Regression Analysis.* New York: Wiley.

Seber, G. A. F. (1984). *Multivariate Observations.* New York: Wiley.

Received January 1991; *revised July and October* 1991; *accepted October* 1991.